UCLA
Department of Statistics Papers

Title
Model Conditions for Asymptotic Robustness in the Analysis of Linear Relations

Permalink
https://escholarship.org/uc/item/4qb329x0

Authors
Satorra, Albert
Bentler, Peter

Publication Date
1990-03-29

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

CAMBRIDGE
UNIVERSITY PRESS

Asymptotic Robustness in Multiple Group Linear-Latent Variable Models
Author(s): Albert Satorra
Source: *Econometric Theory*, Vol. 18, No. 2 (Apr., 2002), pp. 297-312
Published by: Cambridge University Press
Stable URL: http://www.jstor.org/stable/3533294
Accessed: 16/05/2011 18:34

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=cup.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Cambridge University Press is collaborating with JSTOR to digitize, preserve and extend access to *Econometric Theory*.

http://www.jstor.org

Econometric Theory, **18**, 2002, 297–312. Printed in the United States of America.
DOI: 10.1017.S0266466602182041

ASYMPTOTIC ROBUSTNESS IN MULTIPLE GROUP LINEAR-LATENT VARIABLE MODELS

ALBERT SATORRA
Universitat Pompeu Fabra

Standard methods for analyzing linear-latent variable models rely on the assumption that the observed variables are normally distributed. Normality allows statistical inferences to be carried out based solely on the first-and second-order moments. In general, inferences for nonnormally distributed data require the estimates of matrices of third-and fourth-order moments. In the present paper, we show that inferences based on normal theory retain validity and asymptotic efficiency under general assumptions that allow for considerable departure from normality. In particular, we obtain conditions under which correct asymptotic inferences are attained when replacing a matrix of higher order moments by a matrix that depends only on cross-product moments of the data.

1. INTRODUCTION

Mean and covariance structure models are nowadays widely used in social, economic, and behavioral studies to analyze linear relationships among variables, some of which are unobservable (latent) or subject to measurement error. See Abowd and Card (1989), Aasness, Biørn, and Skjerpen (1993), and Behrman, Rosenzweig, and Taubman (1994) for recent applications of these models in econometrics; and, e.g., Anderson (1989) for covariance structure analysis of linear latent variable models.

A common approach to moment structure analysis is based on minimum distance (MD) methods, in which a structured vector $\sigma = \sigma(\vartheta)$ of population moments (usually first- and second-order moments) is fitted to the vector s of corresponding sample moments (Hansen, 1982; Chamberlain, 1982; Browne, 1984; Abowd and Card, 1989). The asymptotic variance matrix Γ_s of $\sqrt{n}s$, where n is sample size, plays a fundamental role in designing an efficient MD analysis and in assessing the sampling variability of the statistics of interest. When the observed variables are normally distributed, Γ_s is a function of first- and second-order moments only; in the general case of nonnormal data, Γ_s in-

I thank T.W. Anderson, Risto D.H. Heijmans, and Nicholas T. Longford for very helpful comments on an earlier draft of this paper. The constructive comments of the co-editor and an anonymous reviewer are very much appreciated. Any remaining errors are of course mine. The work was supported by the Spanish DGICYT grant PB96-0300 and BEC2000-0983. Address correspondence to: Albert Satorra, Universitat Pompeu Fabra, Department of Economics and Business, Ramon Trias Fargas 23, 08005 Barcelona, Spain; e-mail: satorra@upf.es.

© 2002 Cambridge University Press 0266-4666/02 $9.50

volves higher order moments of the observed variables. Here s is the (symmetric) vectorization of an augmented moment matrix, and the form of Γ_s under the normality assumption is denoted as Γ_s^*.

In the present paper we investigate conditions under which the replacement of Γ_s by Γ_s^* (or simply by Ω, a matrix given in Section 3, which is a function of population cross-product moments only) gives asymptotically valid inferences when the data are nonnormal. By using a consistent estimator of Ω instead of Γ_s, we avoid sample third- and fourth-order moments. Estimates of third- and fourth-order moments tend to be highly unstable in small samples, thus giving rise to small sample size distortions in the analysis (for recent investigations on small-sample properties of MD estimates in covariance structures, see, e.g., Altonji and Segal, 1996; and Clark, 1996). See also Horowitz (1998) for an alternative based on bootstrap methods.

The validity of inferences based on the normality assumption when the data are not normally distributed has been called asymptotic robustness (Anderson, 1987). Asymptotic robustness was investigated first in the context of factor analysis models (Anderson and Amemiya, 1988; Amemiya and Anderson, 1990; Browne, 1987) and then extended to a wider class of models and statistics (e.g., Browne and Shapiro, 1988; Anderson, 1989; Satorra and Bentler, 1990; Satorra and Neudecker, 1994).

The present paper extends the work on asymptotic robustness in several aspects. As in Satorra (1993), we consider asymptotic robustness in the general setting of multiple group analysis, where a common model is fitted to several independent samples (or groups), with model parameters possibly restricted to be equal across groups. Such an analysis allows us, among other possibilities, to investigate between-population differences of model parameters and to combine information from different samples. We also adopt Anderson and Amemiya's device of a Taylor series expansion around a sample dependent value of the parameter vector. With this approach, we do not have to insist on finiteness of the matrix Γ_s and we can accommodate a latent component that can be assumed fixed (across hypothetical sample replications). We obtain results for models that structure means and covariances of observable variables, encompassing a wide class of structural equation models that include regression with errors in variables, path analysis models, econometric simultaneous equation models, multivariate regression, models for panel data, etc. We address methods that can be implemented directly in standard computer software such as LISREL of Jöreskog and Sörbom (1989), EQS of Bentler (1989), and the procedure CALIS in SAS (1990).

The paper is organized as follows. Section 2 presents the model setup and describes a specific example. Section 3 describes general MD estimation and the (normal theory) NT approach. Section 4 presents the results on asymptotic robustness. Section 5 concludes. Proofs are confined to the Appendix.

The following notation is used. For a symmetric matrix A, $\mathrm{v}(A)$ is the vector obtained from $\mathrm{vec}(A)$ by eliminating the duplicated elements associated with the symmetry of A; D and D^+ are matrices defined by the identities

$\text{vec}(A) = D\text{v}(A)$ and $\text{v}(A) = D^{+}\text{vec}(A)$, for a symmetric matrix A (Magnus and Neudecker, 1999). We use the standard notation $\bigoplus_{i=1}^{m} A_i$ for the direct sum of matrices (i.e., the block-diagonal matrix with blocks $A_1, A_2, \ldots, A_m$). Given a (vector) statistic a, we denote by Γ_a the asymptotic variance matrix of $\sqrt{n}a$, where n is the sample size associated with a.

2. MODEL SETUP

Let $\{z_{gi} : i = 1, \ldots, n_g; g = 1, \ldots, G\}$ be multiple group data, where z_{gi} is a $p_g \times 1$ vector of observable variables, i indexes individuals, and g indexes groups (note that we allow the dimension of z_{gi} to vary with g). The samples $\{z_{gi}\}_i$ are assumed to be mutually independent across g. Define the $p_g \times p_g$ matrix of (uncentered) sample cross-product moments $S_g = (1/n_g)\sum_{i=1}^{n_g} z_{gi} z'_{gi}$ and let $s = (s'_1, \ldots, s'_G)'$, where $s_g = \text{v}(S_g)$, be the overall $p^* \times 1$ vector of sample moments, where $p^* = \sum_{g=1}^{G} p_g^*$ and $p_g^* = p_g(p_g + 1)/2$. Let Σ_g be the probability limit of S_g as $n_g \to \infty$ and let the $p^* \times 1$ vector $\sigma = \{(\text{v}(\Sigma_1))', \ldots, (\text{v}(\Sigma_G))'\}'$ be the probability limit of s.

Assume the multivariate linear relation

$$
\begin{aligned}
z_{gi} &= \sum_{\ell=0}^{L_g} \Lambda_g^{(\ell)} \xi_{g\ell i} \\
&= \Lambda_g^{(0)} \xi_{g0i} + \sum_{\ell=1}^{L_g - 1} \Lambda_g^{(\ell)} \xi_{g\ell i} + \Lambda_g^{(L_g)} \xi_{gL_g i}, \quad g = 1, \ldots, G, \qquad \textbf{(1)}
\end{aligned}
$$

where the $\Lambda_g^{(\ell)}$ are $p_g \times m_{g\ell}$ matrices of coefficient and the $\xi_{g\ell i}$ are $m_{g\ell} \times 1$ vector variables. The $\xi_{g\ell i}$, $\ell = 0, \ldots, L_g$, are of three types (fixed, distribution free, and normal), as depicted in (1) and implied by the conditions

(a) the $\{\xi_{g0i}\}_i$ are fixed in repeated sampling, with

$$
\lim_{n_g \to +\infty} \frac{1}{n_g} \sum_{i=1}^{n_g} \xi_{g0i} = \mu_{0g} \quad \text{and} \quad \lim_{n_g \to +\infty} \frac{1}{n_g} \sum_{i=1}^{n_g} \xi_{g0i} \xi'_{g0i} = \Phi_g^{(0)},
$$

for suitable $m_{g0} \times 1$ vector μ_{0g} and $m_{g0} \times m_{g0}$ matrix $\Phi_g^{(0)}$

(b) the $\{\xi_{g\ell i}\}_i$, $\ell = 1, \ldots, L_g$, are independent and identically distributed (i.i.d.) sequences of zero mean and (finite) $m_{g\ell} \times m_{g\ell}$ variance matrices, $\Phi_g^{(\ell)}$, with $E(\xi'_{ghi} \xi'_{g\ell i}) = 0$, when $h \neq \ell$ (i.e., uncorrelated).

(c) the $\{\xi_{gL_g i}\}_i$ are i.i.d. normal.

In the models considered in the present paper, the Λ's and Φ's depend on unknown parameters, and the aim is to carry out inferences about these parameters.

The theory of the present paper also applies when the fixed and normal components are absent from (1); when ξ_{g0i} is absent, $E(z_{ig}) = 0$. Without loss of generality, to accommodate models that impose a structure on the means of observable variables in addition to the covariances, we let the first component of z_{gi} and ξ_{g0i} be the constant unit element (the first row of $\Lambda_g^{(0)}$ is then restricted accordingly).

We write (1) in the compact form

$$z_{gi} = \Lambda_g \xi_{gi}, \quad g = 1 \ldots, G, \tag{2}$$

where

$$\Lambda_g = (\Lambda_g^{(0)}, \Lambda_g^{(1)}, \ldots, \Lambda_g^{(L_g)})$$

and

$$\xi_{gi} = (\xi'_{g0i}, \xi'_{g1i}, \ldots, \xi'_{gL_g i})'$$

are, respectively, a $p_g \times m_g$ matrix of coefficients and a $m_g \times 1$ vector variable, with $m_g = \sum_{\ell=0}^{L_g} m_{g\ell}$. From (2), we obtain

$$\Sigma_g = \Lambda_g \Phi_g \Lambda'_g, \quad g = 1, \ldots, G, \tag{3}$$

where Φ_g is the probability limit of

$$Q_g = n_g^{-1} \sum_{i=1}^{n_g} \xi_{gi} \xi'_{gi}. \tag{4}$$

Note that

$$\Phi_g = \bigoplus_{\ell=1}^{L_g} \Phi_g^{(\ell)},$$

as a result of the conditions (a) and (b) given previously.

We assume the model implies (twice continuously differentiable) matrix-valued functions $\Lambda_g = \Lambda_g(\tau)$ and $\Phi_g^{(L_g)} = \Phi_g^{(L_g)}(\tau)$, where τ is a $t^* \times 1$ subvector of the $(t \times 1)$ vector ϑ of model parameters. Thus, using (3), we obtain the multiple group moment structure

$$\Sigma_g = \Sigma_g(\vartheta), \quad g = 1, \ldots, G, \tag{5}$$

where $\Sigma_g(\vartheta) = \Lambda_g(\tau)\Phi_g(\vartheta)\Lambda_g(\tau)'$. The matrix-valued functions $\Phi_g(\vartheta)$ are assumed to be (twice continuously) differentiable. In a given model, ξ_{gi} contains components of z_{gi} (observed variables), unobservable variables such as measurement errors, disturbance terms of regression equations, latent factors, etc. Note that (5) can be expressed as $\sigma = \sigma(\vartheta)$ where $\sigma(\vartheta)$ is (twice continuously) differentiable. We assume the model is identified in the sense that $\theta \neq \theta^*$ implies that $\sigma(\theta) \neq \sigma(\theta^*)$. For further use, we define the Jacobian $p^* \times t$ matrix $\Delta = \partial\sigma(\vartheta)/\partial\vartheta'$.

We now give a concrete example of this general model setup.

2.1. Model Example: Regression with Errors in Variables

Consider the two-group regression model

$$y^*_{gi} = \alpha + \beta x_{gi} + v_{gi}, \quad i = 1, \ldots, n_g, \tag{6}$$

where for case i in group g ($g = 1,2$), y^*_{gi} and x_{gi} are the values of the response and explanatory variables, respectively, v_{gi} is the value of the disturbance term, α is the intercept, and β is the regression coefficient. Instead of x_{gi} we observe two variables x^*_{1gi} and x^*_{2gi} related to x_{gi} by the following measurement-error equations:

$$\begin{cases} x^*_{1gi} & = & x_{gi} + u_{g1i} \\ x^*_{2gi} & = & x_{gi} + u_{g2i} \end{cases}, \tag{7}$$

where u_{g1i}, u_{g2i}, v_{gi}, and x_{gi} are uncorrelated variables. Assume a common value for the variances of u_{g1i} and u_{g2i}, σ^2_u (the same across groups) and group varying variances for v_{gi} and x_{gi}, $\sigma^2_{v_g}$ and $\sigma^2_{x_g}$, respectively. When the latent regressor is fixed (the so-called fixed x case), we assume the limits $\lim_{n_g \to \infty} n_g^{-1} \sum_{i=1}^{n_g} x^2_{gi}$, $g = 1,2$, are finite.

The model given by equations (6) and (7) and associated assumptions is identified (for a comprehensive overview of measurement-error models in regression analysis, see Fuller, 1987). The analysis of this type of model is usually carried out under the assumption that the observable variables are normally distributed; in the present paper, however, we are concerned with the validity of the NT approach when the latent regressor x_{gi} and the disturbance terms v_{gi} deviate from the normality assumption. This model has the form of (2) when we set

$$z_{gi} = \begin{pmatrix} 1 \\ y^*_{gi} \\ x^*_{1gi} \\ x^*_{2gi} \end{pmatrix}, \quad \xi_{gi} = \begin{pmatrix} 1 \\ x_{gi} \\ v_{gi} \\ u_{1gi} \\ u_{2gi} \end{pmatrix}, \quad \Lambda_g = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \alpha & \beta & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

and

$$\Phi_g = E(\xi_{gi}\xi'_{gi}) = \mathrm{diag}(1, \sigma^2_{x_g}, \sigma^2_{v_g}, \sigma^2_u, \sigma^2_u), \tag{8}$$

with

$$\vartheta = (\alpha, \beta, \sigma^2_u, \sigma^2_{x_1}, \sigma^2_{v_1}, \sigma^2_{x_2}, \sigma^2_{v_2})'.$$

In the "fixed-x" case, ξ_{g0i} is $(1, x'_{gi})'$ and the nonnormal components are just v_{gi}. When the latent regressor is random, ξ_{g0i} is the constant unit element and the nonnormal components are v_{gi} and x_{gi}; in both cases, $\tau = (\alpha, \beta, \sigma^2_u)'$, and the normal component $\xi_{gL_g i}$ is $(u_{1g}, u_{2g})'$.

The issue we address in the present paper concerns the analysis of the models described previously using MD methods (or equivalent maximum likelihood [ML] methods) that are based on the assumption that the z_{gi} are i.i.d. normally distributed. It will be seen that NT statistics produces correct inferences despite violation of the normality assumption. For this example, it will be seen that NT gives correct inferences for estimates of τ and for the chi-square goodness-of-fit test of the model, despite severe deviation from normality of x_{gi} and v_{gi}.

3. MD ANALYSIS

Consider the MD estimator

$$\hat{\vartheta} = \operatorname{argmin}_{\vartheta \in \Theta}\{s - \sigma(\vartheta)\}'\hat{V}\{s - \sigma(\vartheta)\}, \tag{9}$$

where $\hat{V}$ is a stochastic $p^* \times p^*$ matrix with $\hat{V} \xrightarrow{P} V$, a positive definite matrix, and Θ is a t-dimensional compact subset of $\mathcal{R}^t$. The estimator $\hat{\vartheta}$ is a generalized method of moments (GMM) estimator (Hansen, 1982). Assume that $\hat{\vartheta}$ in (9) is unique for all s, the $t \times t$ matrix $\Delta' V \Delta$ is nonsingular, and the true parameter value ϑ is in the interior of Θ.

In this setup, the standard asymptotic theory of MD estimation implies (see, e.g., Chamberlain, 1982) that $\hat{\vartheta}$ is consistent and asymptotically normal with variance matrix

$$\operatorname{avar}(\hat{\vartheta} | V, \Gamma_s) = \frac{1}{n}(\Delta' V \Delta)^{-1}\Delta' V \Gamma_s V \Delta(\Delta' V \Delta)^{-1}, \tag{10}$$

where $n = n_1 + \ldots + n_G$ is the overall sample size and $\Gamma_s = \operatorname{avar}(\sqrt{n}s)$ is the $p^* \times p^*$ asymptotic variance matrix of the sample moments. When $V\Gamma_s V = V$ (alternatively, when $\Gamma_s V \Gamma_s = \Gamma_s$ and Δ is in the column space of Γ_s) then (10) simplifies to

$$\operatorname{avar}(\hat{\vartheta} | V) = \frac{1}{n}(\Delta' V \Delta)^{-1}. \tag{11}$$

In this case, the corresponding MD estimator is asymptotically optimal within the class of MD estimators based on s (Hansen, 1982). When V is an identity matrix, we have the equally weighted MD estimator.

Because the vectors s_g are uncorrelated, we have $\Gamma_s = \bigoplus_{g=1}^{G} \pi_g^{-1}\Gamma_{s_g}$, with $\pi_g = \lim_{n\to\infty}(n_g/n)$. Let assume further that $\pi_g > 0$ for all g and that $s - \sigma(\vartheta)$ and Δ are in the column space of Γ_s, for all ϑ and s. Under standard regularity conditions, the fourth-order moment matrix

$$\hat{\Gamma}_{s_g} = \frac{1}{n_g - 1}\sum_{i=1}^{n_g}(d_{gi} - s_g)(d_{gi} - s_g)',$$

where $d_{gi} = \mathrm{v}(z_{gi}z_{gi}')$, is a consistent estimator of Γ_{s_g}, and so

$$\hat{\Gamma}_s = \bigoplus_{g=1}^{G}\frac{n}{n_g}\hat{\Gamma}_{s_g}, \tag{12}$$

is a consistent estimator of Γ_s. The matrix $\hat{\Gamma}_s$ is called the asymptotic robust (AR) estimator of Γ_s, to indicate that no specific distributional assumption is used. When in (10) we replace Γ_s, V, and Δ by their respective estimators $\hat{\Gamma}_s$, $\hat{V}$, and $\hat{\Delta}$, we obtain the so-called AR estimator of the variance matrix of $\hat{\vartheta}$.

A goodness-of-fit test statistic for the overall model is

$$T_V = n(s - \hat{\sigma})'\hat{\Delta}_\perp(\hat{\Delta}_\perp'\hat{\Gamma}_s\hat{\Delta}_\perp)^+\hat{\Delta}_\perp'(s - \hat{\sigma}), \tag{13}$$

where $\hat{\Gamma}_s$ is the AR estimator of Γ, $\hat{\sigma} = \sigma(\hat{\vartheta})$, $\hat{\vartheta}$ is the MD estimator of ϑ, and $\hat{\Delta}_\perp$ is a consistent estimator of an orthogonal complement $\Delta_\perp$ of Δ (i.e., $\Delta_\perp$ is a $p^* \times (p^* - G - t)$ matrix of full column rank, such that $\Delta_\perp'\Delta = 0$). Note that Δ has G rows identical to zero as a result of the constant unit element of z_{gi} (which gives a structural 1 in each vector s_g).

When we assume finiteness of the matrix Γ_s, then T_V is asymptotically chi-square distributed with

$$r = \text{rank}\{\Delta_\perp'\Gamma_s\Delta_\perp\} \tag{14}$$

degrees of freedom (Browne, 1984; Newey, 1985). We call T_V the AR goodness-of-fit test statistic.

Because estimators of higher order moments tend to be highly unstable in small samples, the use of $\hat{\Gamma}_s$ may lead to small sample size distortion of inferences, especially when $\hat{\Gamma}_s$ is involved in obtaining the optimal MD estimator. See Horowitz (1998) for a discussion of the small sample size distortions of the optimal MD estimator, with a proposal based on bootstrap methods for improving inferences in MD analysis of covariance structures. Third- and fourth-order sample moments are of course avoided when we assume that z_{gi} is normally distributed. It is a matter of theoretical and practical interest whether the approach based on normality remains valid with nonnormal data. Section 4, which follows, provides a theorem with results on this issue. First, however, we need to describe the NT approach.

3.1. The NT Approach

When the $\{z_{gi}\}$ are assumed to be i.i.d. normal, i.e., under the NT assumption, then Γ_s equals (see, e.g., Satorra, 1993) $\Gamma_s = \Gamma_s^* = \Omega - \Upsilon$, with

$$\Omega = \bigoplus_{g=1}^G \frac{1}{\pi_g}\Omega_g, \quad \text{with} \quad \Omega_g = 2D^+(\Sigma_g \otimes \Sigma_g)D^{+\prime},$$

and

$$\Upsilon = \bigoplus_{g=1}^G \frac{1}{\pi_g}\Upsilon_g, \quad \text{with} \quad \Upsilon_g = 2D^+(\mu_g\mu_g' \otimes \mu_g\mu_g')D^{+\prime},$$

and $\mu_g = E(z_{gi})$, a $p_g \times 1$ vector.

To define the NT-MD estimator, consider

$$V^* = \bigoplus_{g=1}^G \pi_g V_g^*, \quad \text{with} \quad V_g^* = \frac{1}{2}D'(\Sigma_g^{-1} \otimes \Sigma_g^{-1})D$$

and let $\hat{\Omega}$ and $\hat{V}^*$ be the respective matrices Ω and V^* with π_g and Σ_g substituted by n_g/n and S_g, respectively. MD estimation with $V = V^*$ will be denoted as NT-MD. Note that $V^* = \Omega^{-1}$. Because V_g^* is a g-inverse of Γ_g^* (Satorra and Neudecker, 1993), we have that V^* is a g-inverse of Γ^*,

$$\Gamma^* V^* \Gamma^* = \Gamma^*. \tag{15}$$

Under NT, the asymptotic variance matrix (10) of the NT-MD estimator reduces to (11) (we used that Δ is in the column space of Γ^* and (15)). The standard errors extracted from (11) will be called the NT standard errors.

When NT holds and the z_{gi} have zero mean, the goodness-of-fit test T_V of (13) is asymptotically equivalent to

$$T_V^* = n(s - \hat{\sigma})'\hat{\Delta}_\perp(\hat{\Delta}_\perp'\hat{\Omega}\hat{\Delta}_\perp)^+\hat{\Delta}_\perp'(s - \hat{\sigma}) \tag{16}$$

($\hat{\Gamma}_s$ is just replaced by $\hat{\Omega}$) and, because $\hat{\Omega}$ is nonsingular,

$$T_V^* = n(s - \hat{\sigma})'\{\hat{\Omega}^{-1} - \hat{\Omega}^{-1}\hat{\Delta}(\hat{\Delta}'\hat{\Omega}^{-1}\hat{\Delta})^{-1}\hat{\Delta}'\hat{\Omega}^{-1}\}(s - \hat{\sigma}).$$

An alternative approach to NT-MD estimation is pseudo-maximum likelihood (PML), where the function to be minimized is an affine transformation of the log-likelihood function (under NT),

$$F\{s, \sigma(\vartheta)\} = \sum_{g=1}^{G} \frac{n_g}{n} [\log|\Sigma_g(\vartheta)| + \mathrm{tr}\{S_g \Sigma_g(\vartheta)^{-1}\} - \log|S_g| - p_g]. \tag{17}$$

The minimizer of $F\{s, \sigma(\vartheta)\}$ is a PML estimator (a closest moments estimator, in the notation of Newey, 1988). We note that the PML estimator is asymptotically equivalent to the NT-MD estimator, because the Hessian matrix $(\partial^2/\partial\sigma\partial\sigma')(\frac{1}{2})F(s,\sigma)$ evaluated at $(s,\sigma) = (\sigma,\sigma)$ equals V^* (Shapiro, 1986; Newey, 1986). An alternative PML goodness-of-fit test statistic is $n\hat{F} = nF\{s, \sigma(\hat{\vartheta}_{PML})\}$, where $\hat{\vartheta}_{PML}$ is the PML estimator and $n\hat{F}$ has the same asymptotic distribution as $T_{V^*}^*$.

4. ASYMPTOTIC ROBUSTNESS

4.1. Key Assumptions

In this section we develop the assumptions needed for asymptotic robustness. First, however, we need to introduce basic matrices and vectors of sample and population moments.

The matrix Q_g of (4) has matrix blocks $Q_g^{(rt)} = n_g^{-1}\sum_{i=1}^{n_g} \xi_{gri}\xi_{gti}'$ $(m_{gr} \times m_{gt})$, $0 \le r,t \le L_g$. Set $q = \{(\mathrm{v}(Q_1))', \ldots, (\mathrm{v}(Q_G))'\}'$, a $m^* \times 1$ vector, where $m^* = \sum_{g=1}^{G} m_g^*$ and $m_g^* = m_g(m_g + 1)/2$, and let $s = \Xi q$, where $\Xi = \bigoplus_{g=1}^{G} \Xi_g$, a $p^* \times m^*$ matrix, with $\Xi_g = D^+(\Lambda_g \otimes \Lambda_g)D$ of dimension $p_g^* \times m_g^*$. We decompose Q_g as $Q_g = \bar{Q}_g + \tilde{Q}_g$, with

$$\tilde{Q}_g = \begin{pmatrix} \bigoplus_{\ell=0}^{L_g-1} Q_g^{(\ell)} & 0 \\ 0 & 0_{m_{L_g}\times m_{L_g}} \end{pmatrix} \tag{18}$$

($Q_g^{(\ell)} = Q_g^{(\ell\ell)}$ and m_{L_g} is the dimension of ξ_{L_g}) so that

$$s = \Xi\bar{q} + \Xi\tilde{q}, \tag{19}$$

where $\tilde{q} = \{(\mathrm{v}(\tilde{Q}_1))', \ldots, (\mathrm{v}(\tilde{Q}_G))'\}'$ and $\bar{q} = q - \tilde{q}$. Note that $\bar{q}$ involves no elements of the matrices $Q_g^{(\ell)}$, $g = 1, \ldots G$, $\ell = 0, \ldots, L_{g-1}$. Further, we can rewrite (19) as

$$s = \Xi\bar{q} + \Xi_u\tilde{q}_u, \tag{20}$$

where $\Xi_u = \Xi R$ and $\tilde{q}_u = Rq$ is the $m_u^* \times 1$ vector $\tilde{q}$ pruned from its structural $0 - 1$ elements, with R an $m^* \times m_u^*$ matrix of 0's and 1's. From (18) and the presence of the unit component in ξ_{g0i}, we have $m_u^* = \sum_{g=1}^{G} \sum_{\ell=0}^{L_g-1} (m_{g\ell}^* - 1)$.

By taking the probability limit of (20), we obtain

$$\sigma = \Xi\bar{\phi} + \Xi_u\tilde{\phi}_u, \tag{21}$$

where $\bar{\phi}$ and $\tilde{\phi}_u$ are the probability limits of $\bar{q}$ and $\tilde{q}_u$, respectively. Because $\xi_{i\ell g}$, $\ell = 1, \ldots, L_g$, are assumed to be uncorrelated and of zero mean, it holds that $Q_g^{(ht)} \xrightarrow{P} 0$ when $h \neq t$, and thus $\bar{\phi}$ collects only structural 0's and the nonredundant elements of the matrices $\Phi_g^{(L_g)}$, $g = 1, \ldots, G$.

The following assumption ensures the elements of $\mathrm{v}(\Phi_{g\ell})$ ($\ell = 1, \ldots, L_g - 1$, $g = 1, \ldots, G$) are unrestricted parameters of the model.

Assumption A.1 (Model Assumption, MA). The elements of $\tilde{\phi}_u$ are free parameters of the model (i.e., we can write $\vartheta = (\tau', \tilde{\phi}_u')'$).

From (21), A.1 implies

$$\sigma = \sigma(\vartheta) = \Xi(\tau)\bar{\phi}(\tau) + \Xi_u(\tau)\tilde{\phi}_u, \tag{22}$$

with $\sigma(.)$ a (twice continuously differentiable) function of ϑ. This parameterization of σ yields the following partition:

$$\Delta = [\Delta_\tau, \Xi_u], \tag{23}$$

with $\Delta_\tau = \partial\sigma(\vartheta)/\partial\tau'$. It will be seen that this partition of Δ implies that the asymptotic distribution of statistics of interest (estimates of τ and goodness-of-fit test statistics) depends only on the asymptotic distribution of $\bar{q}$ of (20) and not on $\tilde{q}_u$.

The next assumption imposes mutual independence (not only no correlation) among random constituents of the model. It will be required to ensure that the asymptotic distribution of $\bar{q}$ involves no higher order moments of the z_{gi} (see Lemma 1 in the Appendix).

Assumption A.2 (Independence Assumption, IA). The components ξ_{g1i}, $\xi_{g2i}, \ldots, \xi_{gL_gi}$ are mutually independent (not only uncorrelated).

This assumption may be restrictive in some applications. For example, in a regression model, statistical independence (not just no correlation) between regressors and the disturbance term does not allow for heteroskedasticity.

The following assumption of zero skewness will be needed to obtain results of asymptotic efficiency.

Assumption A.3 (Zero Skewness, ZS). For each sample g and $\ell > 0$,

$$E(\xi_{g\ell i}\xi'_{g\ell i}(\xi_{g\ell i})_j) = 0$$

for every j. (That is, all third-order moments vanish.)

4.2. Main Results

From (19), we have

$$\Gamma_s = \Xi\Gamma_{\bar{q}}\Xi' + \Xi_u\Gamma_{\tilde{q}_u}\Xi'_u + \Xi\Gamma_{\bar{q},\tilde{q}_u}\Xi'_u + \Xi_u\Gamma_{\tilde{q}_u,\bar{q}}\Xi', \tag{24}$$

where $\Gamma_{\bar{q}}$ and $\Gamma_{\tilde{q}_u}$ denote, respectively, the asymptotic variance matrices of $\bar{q}$ and $\tilde{q}_u$ and $\Gamma_{\bar{q},\tilde{q}_u}$ and $\Gamma_{\tilde{q}_u,\bar{q}}$ denote matrices of asymptotic covariances. Let J be the $t^* \times t$ selection matrix such that $\hat{\tau} = J\hat{\vartheta}$.

When A.1 holds, the partition (23) of Δ applies; thus

$$\Delta'_\perp \Xi_u = 0 \tag{25}$$

and

$$J(\Delta' V\Delta)^{-1}\Delta' V\Xi_u = 0; \tag{26}$$

consequently, the matrix $\Delta_\perp\Gamma_s\Delta'_\perp$, involved in the expression of the goodness-of-fit test T_V of (13) and the leading $t^* \times t^*$ principal submatrix of the estimate's variance matrix (26), is free of the terms on the right-hand side of (24) except for the first term. That is, the asymptotic distribution of T_V and $\hat{\tau}$ is determined only by the distribution of $\bar{q}$ (i.e., is free of the distribution of $\tilde{q}_u$).

On the other hand, when A.2 holds, the matrix $\Gamma_{\bar{q}}$ is free of the higher order moments of the variables $\xi_{i\ell g}$, $\ell = 1,\dots,L_g - 1$ (see Lemma 1 in the Appendix). Thus, combining A.1 and A.2 we obtain that T_V^* and $\hat{\tau}$ are free (asymptotically) of higher order moments of the data. This is made precise in the following theorem, a detailed proof of which appears in the Appendix.

THEOREM 1. *Under Assumptions A.1 and A.2,*

1. $\sqrt{n}(\hat{\tau} - \tau)$ *is asymptotically normally distributed with* $\operatorname{avar}(\hat{\tau}) = \operatorname{avar}(\hat{\tau}|V,\Gamma_s^*)$ *(i.e., the corresponding submatrix of (10) with* $\Gamma_s = \Gamma_s^*$*);*
2. *when* $V = V^*$, $\operatorname{avar}(\hat{\tau})$ *is the* $t^* \times t^*$ *upper left-hand corner submatrix of* $\Delta' V^*\Delta$ *and* $\hat{\tau}$ *is the minimum variance estimator (in the class of MD estimators of (9), for any* V*);*
3. *when additionally A.3 holds and* $V = V^*$*, the whole vector of estimators* $\hat{\vartheta}$ *has asymptotic minimum variance (in the class of MD estimators of (9), for any* V*);*

4. *$\sqrt{n}(s - \hat{\sigma})$ is asymptotically normal, with zero mean and variance matrix determined by τ, V, and Γ^* (i.e., the asymptotic distribution $\sqrt{n}(s - \hat{\sigma})$ is free of higher order moments of the z_{gi});*
5. *the asymptotic distribution of the goodness-of-fit test statistic T_V^* of (16) (for any V) is chi square with degrees of freedom given by (14).*

Proof. See the Appendix.

Note that normality is not assumed and that A.3 is needed only for result 3 of the theorem. Note that except for results 2 and 3, the theorem holds for any MD estimation method (i.e., for any matrix V used).

In many cases the model is not exactly true; i.e., it is only an approximate model. In this circumstance, to ensure finite asymptotic distribution of test statistics, we proceed as follows. On the right-hand side of the multivariate relation (1), we add the misspecification term ν_{gi}, with the condition that for all g, ℓ, $\lim_{n_g \to \infty}(1/\sqrt{n_g})\sum_{i=1}^{n_g} \xi_{g\ell i}\nu'_{gi}$ and $\lim_{n_g \to \infty}(1/\sqrt{n_g})\sum_{i=1}^{n_g} \nu_{gi}\nu'_{gi}$ are finite, possibly zero. The presence of such misspecification terms implies a population drift assumption, $\sigma^0 = \sigma_n^0$, for the probability limit of s, σ^0, with the property that $\delta = \lim_{n\to\infty}\sqrt{n}(\sigma_n^0 - \sigma_0)$ is a finite $p^* \times 1$ vector, σ_0 being the probability limit of $\hat{\sigma}$. This leads to the classical device of a sequence of local alternatives used to investigate the asymptotic distribution of test statistics when the model does not hold exactly (cf. Stroud, 1972). Because this parameter drift assumption implies that the "size" of the misspecification decreases to zero as sample size $n \to \infty$, for the asymptotic approximations to be accurate in finite samples, large n but "small" misspecification errors are required.

Under this parameter drift assumption, it is easy to see that Theorem 1 applies with the only modification on results 4 and 5, where now $\sqrt{n}(s - \hat{\sigma})$ has mean δ and T_V^* has a noncentral chi-square distributed with noncentrality parameter

$$\lambda = n\delta'\Delta_\perp(\Delta'_\perp\Omega\Delta_\perp)^+\Delta'_\perp\delta. \tag{27}$$

Note that when $\delta = 0$, i.e., when the model is "exact," then $\lambda = 0$ and the central case results are recovered.

5. CONCLUSIONS

We have shown that for a wide variety of linear-latent variable models, the model and independence assumptions (A.1 and A.2, respectively) are sufficient conditions for the asymptotic efficiency of the NT-MD and PML estimators of τ and also for correctness of the associated NT standard errors, despite nonnormality of the data. Furthermore, when the distribution of each latent random constituent of the model is symmetric (i.e., A.3 holds), then asymptotic efficiency applies to all the components of the NT-MD and PML estimators. In addition, the distribution of the residual vector $s - \hat{\sigma}$ and of the goodness-of-fit test statistic T_V^* carry over from NT to the general case; in particular, the NT and PML goodness-of-fit test statistics, $T_{V^*}^*$ and $n\hat{F}$, are asymptotically chi-square distributed when the model holds, despite nonnormality. The practical implications of these findings are that when there is independence among random constituents of the model, and the model does not restrict the variances and covariances of nonnormal constituents (not restricted even by equality across groups), we can avoid using the optimal MD approach (which requires a weight matrix based on higher order moments of the data); instead, we can just use NT-MD or PML methods with the associated NT statistics. This may be useful in view of the small sample size distortions reported recently for the optimal MD estimator (e.g., Altonji and Segal, 1996; Clark, 1996).

We also showed that Γ_s^* can be replaced by Ω; the additional advantage of such a replacement is that standard software for covariance structure analysis (e.g., LISREL of Jöreskog and Sörbom, 1989; EQS of Bentler, 1989) can be used without modification to analyze mean and covariance structures. We have obtained results for NT-MD and PML estimators and also for MD estimators with general weight matrix V (e.g., $V = I$). Moreover, Theorem 1 can easily be extended to the case of a Fisher-consistent estimator that (asymptotically) is a smooth function of s, as, e.g., in instrumental variable estimation.

In the example given in Section 2 of regression with errors in variables, our results ensure correctness of the NT standard errors for estimates of the subvector $\tau = (\alpha, \beta, \sigma_{uu})'$ of ϑ and asymptotic correctness also of the usual NT chi-square goodness-of-fit test statistic, provided there is mutual independence among x_{gi}, v_{gi}, and $u_{gi} = (u_{g1i}, u_{g2i})'$ and u_{gi} is normally distributed; further, this holds for both x_{gi} random or fixed in repeated sampling.

REFERENCES

Aasness, J., E. Biørn, & T. Skjerpen (1993) Engel functions, panel data, and latent variables. *Econometrica* 61, 1395–1422.

Abowd, J.M. & D. Card (1989) On the covariance structure of earnings and hours changes. *Econometrica* 57, 411–445.

Altonji, J.G. & L.M. Segal (1996) Small-sample bias in GMM estimation of covariance structures. *Journal of Business and Economic Statistics* 14, 353–366.

Amemiya, Y. & T.W. Anderson (1990) Asymptotic chi-square tests for a large class of factor analysis models. *Annals of Statistics* 3, 1453–1463.

Anderson, T.W. (1987) Multivariate linear relations. In T. Pukkila & S. Puntanen (eds.), *Proceedings of the Second International Conference in Statistics*, pp. 9–36. Tampere, Finland: University of Tampere.

Anderson, T.W. (1989) Linear latent variable models and covariance structures. *Journal of Econometrics* 41, 91–119.

Anderson, T.W. & Y. Amemiya (1988) The asymptotic normal distribution of estimates in factor analysis under general conditions. *Annals of Statistics* 16, 759–771.

Behrman, J., M. Rosenzweig, & P. Taubman (1994) Endowments and the allocation of schooling in the family and in the marriage market: The twins experiment. *Journal of Political Economy* 102, 1131–1174.

Bentler, P.M. (1989) *EQS Structural Equations Program Manual.* Los Angeles: BMDP Statistical Software.

Browne, M.W. (1984) Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology* 37, 62–83.

Browne, M.W. (1987) Robustness in statistical inference in factor analysis and related models. *Biometrika* 74, 375–384.

Browne, M.W. & A. Shapiro (1988) Robustness of normal theory methods in the analysis of linear latent variable models. *British Journal of Mathematical and Statistical Psychology* 41, 193–208.
Chamberlain, G. (1982) Multivariate regression models for panel data. *Journal of Econometrics* 18, 5–46.
Clark, T.E. (1996) Small-sample properties of estimators of nonlinear models of covariance structure. *Journal of Business and Economic Statistics* 14, 367–373.
Fuller, W.A. (1987) *Measurement Error Models*. New York: Wiley.
Hansen, L.P. (1982) Large sample properties of generalized method of moments estimators. *Econometric* 50, 1029–1054.
Horowitz, J.L. (1998) Bootstrap methods for covariance structures. *Journal of Human Resources* 33, 39–61.
Jöreskog, K. & D. Sörbom (1989) *LISREL 7: A Guide to the Program and Applications*, 2nd ed. Chicago: SPSS.
Magnus J. & H. Neudecker (1999) *Matrix Differential Calculus with Applications in Statistics and Econometrics*, rev. ed. New York: Wiley.
Moore, D.S. (1977) Generalized inverses, Wald's method, and the construction of chi-squared tests of fit. *Journal of the American Statistical Association* 72, 131–137.
Newey, W.K. (1985) Generalized method of moment specification testing. *Journal of Econometrics* 29, 229–256.
Newey, W.K. (1988) Asymptotic equivalence of closest moments and GMM estimators. *Econometric Theory* 4, 336–340.
SAS (1990) CALIS and LOGISTIC Procedures. SAS Technical report P-200, SAS Institute, Cary, North Carolina.
Satorra, A. (1993) Asymptotic robust inferences in multiple-group analysis of augmented-moment structures. In C.M. Cuadras and C.R. Rao (eds.), *Multivariate Analysis: Future Directions 2*, pp. 211–229. Amsterdam: Elsevier.
Satorra, A. & P.M. Bentler (1990) Model conditions for asymptotic robustness in the analysis of linear relations. *Computational Statistics and Data Analysis* 10, 235–249.
Satorra, A. & H. Neudecker (1993) A matrix equality applicable in the analysis of mean and covariance structures. *Econometric Theory* 8, 581.
Satorra, A. & H. Neudecker (1994) On the asymptotic optimality of alternative minimum-distance estimates in linear latent-variable models. *Econometric Theory* 10, 867–883.
Shapiro, A. (1986) Asymptotic equivalence of minimum discrepancy function estimators to G.L.S. estimators. *South African Statistical Journal* 19, 73–81.
Stroud, T.W.F. (1972) Fixed alternatives and Wald's formulation of the noncentral asymptotic behavior of the likelihood ratio statistic. *Annals of Mathematical Statistics* 43, 447–454.

APPENDIX: PROOF OF THEOREM 1

This Appendix provides the proofs of Lemma 1 and Theorem 1.

LEMMA 1.

1. *Under Assumptions A.1 and A.2,*

$$(\tilde{q}_u - \tilde{\phi}_u) = o_p(1); \tag{A.1}$$

$$\sqrt{n}(\bar{q} - \bar{\phi}) \xrightarrow{L} N(0, \Gamma_{\bar{q}}^*), \tag{A.2}$$

where $\Gamma_{\bar{q}}^*$ *denotes the asymptotic variance of* $\bar{q}$ *under the* NT *assumption.*

2. *When in addition to Assumptions A.2 and A.1, Assumption A.3 holds, then $\bar{q}$ and $\tilde{q}_u$ are asymptotically uncorrelated; thus*

$$\Gamma_s^* = \Xi \Gamma_{\bar{q}}^* \Xi' + \Xi_u \Gamma_{\tilde{q}_u}^* \Xi_u'. \tag{A.3}$$

Proof. The result (A.1) follows by applying the law of large numbers element-wise to $\tilde{q}$.

By applying the Lindeberg–Feller version of the central limit theorem to the linear combinations of $\bar{q}$, we obtain the asymptotic normality stated in (A.2) but with a matrix of asymptotic variances $\Gamma_{\bar{q}}$ that is not necessarily equal to the NT one.

To prove the NT form of the variance matrix of (A.2), note that a nonzero element of $\Gamma_{\bar{q}}$ is the asymptotic covariance $\Gamma_{grt,r't'}$ of sample moments $\bar{q}_{grt,\alpha} = (1/n_g)\sum_{i=1}^{n_g}(\xi_{gri})_{\alpha_1}(\xi_{gti})_{\alpha_2}$, and $\bar{q}_{gr't',\alpha'} = (1/n_g)\sum_{i=1}^{n_g}(\xi_{gri})_{\alpha_1'}(\xi_{gti})_{\alpha_2'}$ with $0 \leq r, t, r', t' \leq L_g$, $r \neq t$ unless $r = t = L_g$, and $r' \neq t'$ unless $r' = t' = L_g$. As a result of A.2, $\Gamma_{gL_gL_g,L_gL_g}$ takes the same value as under NT. Because for $r > 0$, $\{\xi_{gri}\}$ are independent zero mean variables,

$$\Gamma_{grt,r't'} = \frac{1}{\pi_g} \lim_{n\to+\infty} \frac{1}{n_g} \sum_{i=1}^{n_g} E\{(\xi_{gri})_{\alpha_1}(\xi_{gti})_{\alpha_2}(\xi_{gr'i})_{\alpha_1'}(\xi_{gt'i})_{\alpha_2'}\}; \tag{A.4}$$

and, by the mutual independence condition of Assumption A.2, $E\{(\xi_{gri})_{\alpha_1}(\xi_{gti})_{\alpha_2}(\xi_{gr'i})_{\alpha_1'}(\xi_{gt'i})_{\alpha_2'}\} = 0$, except when

$r = r' = L_g, \quad t = t' \neq L_g$: then $\Gamma_{grt,r't'} = (\phi^{(L_g)})_{\alpha_1\alpha_1'}(\phi^{(tt)})_{\alpha_2\alpha_2'}$;

$r = r' = 0, \quad t = t' \neq 0$: then $\Gamma_{grt,r't'} = (\phi^{(0)})_{\alpha_1\alpha_1'}(\phi^{(t)})_{\alpha_2\alpha_2'}$;

$r = r', \quad t = t', \quad r \neq t, \quad 0 < r, t < L_g$: then $\Gamma_{grt,r't'} = (\phi^{(r)})_{\alpha_1\alpha_1'}(\phi^{(t)})_{\alpha_2\alpha_2'}$;

in each case, the same values as under NT. This concludes the proof of (A.2).

To prove (A.3), we need to show that $\Gamma_{(\bar{q})_\alpha,(\tilde{q})_{\alpha'}} = 0$. When $(\bar{q})_\alpha$ and $(\tilde{q})_{\alpha'}$ correspond to the same group g, then

$$\Gamma_{(\bar{q})_\alpha,(\tilde{q})_{\alpha'}} = \frac{1}{\pi_g} \lim_{n\to+\infty} \frac{1}{n_g} \sum_{i=1}^{n_g} \operatorname{cov}\{(\xi_{gri})_{\alpha_1}(\xi_{gti})_{\alpha_2}, (\xi_{gr'i})_{\alpha_1'}(\xi_{gr'i})_{\alpha_2'}\},$$

in analogy with (A.4). When $r, t < L_g$, then

$$\operatorname{cov}\{(\xi_{gri})_{\alpha_1}(\xi_{gti})_{\alpha_2}, (\xi_{gr'i})_{\alpha_1'}(\xi_{gr'i})_{\alpha_2'}\} = E\{(\xi_{gri})_{\alpha_1}(\xi_{gti})_{\alpha_2}(\xi_{gr'i})_{\alpha_1'}(\xi_{gr'i})_{\alpha_2'}\},$$

which equals zero in all cases except when $t = 0$ and $r = r' \neq 0$, in which case

$$= \frac{1}{\pi_g} \lim_{n\to+\infty} \frac{1}{n_g} \sum_{i=1}^{n_g} E\{(\xi_{gri})_{\alpha_1}(\xi_{gti})_{\alpha_2}(\xi_{gr'i})_{\alpha_1'}(\xi_{gr'i})_{\alpha_2'}\}$$

$$= \frac{1}{\pi_g} E\{(\xi_{gri})_{\alpha_1}(\xi_{gri})_{\alpha_1'}(\xi_{gri})_{\alpha_2}\} \left\{ \lim_{n\to+\infty} \frac{1}{n_g} \sum_{i=1}^{n_g} (\xi_{g0i})_{\alpha_2} \right\}$$

$$= \frac{1}{\pi_g} (\nu_g)_{\alpha_2} E\{(\xi_{gri})_{\alpha_1}(\xi_{gri})_{\alpha_1'}(\xi_{gri})_{\alpha_2}\} = 0,$$

by virtue of Assumption A.3. Now, when $r = t = L_g$,

$$\text{cov}\{(\xi_{gri})_{\alpha_1}(\xi_{gti})_{\alpha_2}, (\xi_{gr'i})_{\alpha'_1}(\xi_{gr'i})_{\alpha'_2}\} = E(AB) - E(A)E(B) = 0,$$

where $A = (\xi_{gri})_{\alpha_1}(\xi_{gti})_{\alpha_2}$ and $B = (\xi_{gr'i})_{\alpha'_1}(\xi_{gr'i})_{\alpha'_2}$, because, when $r' \neq L_g$, $(\xi_{gri})_{\alpha_1}(\xi_{gti})_{\alpha_2}$ is independent of $(\xi_{gr'i})_{\alpha'_1}(\xi_{gr'i})_{\alpha'_2}$ and hence $E(AB) = E(A)E(B)$. ■

We prove the statements of Theorem 1 one at a time. Because $s \xrightarrow{P} \sigma$, by standard theory of MD estimation also $\hat{\vartheta} \xrightarrow{P} \vartheta$. Following Anderson and Amemiya (1988) and Amemiya and Anderson (1990) we consider the "mixture" parameter vector $\tilde{\vartheta} = (\tau', \tilde{q}'_u)'$ in which the elements of τ are population values and the elements of $\tilde{q}_u$ are sample specific, and we put $\tilde{\sigma} = \sigma(\tilde{\vartheta})$. Note that by Lemma 1, $\tilde{\vartheta} \xrightarrow{P} \vartheta$. Using (20), (22), and (A.2), we obtain

$$\sqrt{n}(s - \tilde{\sigma}) = \Xi\sqrt{n}(\bar{q} - \bar{\phi}) = O_p(1). \qquad \textbf{(A.5)}$$

By definition of an MD estimator,

$$\hat{\Delta}'\hat{V}(s - \hat{\sigma}) = 0; \qquad \textbf{(A.6)}$$

thus, a Taylor series expansion of $\Delta' V\{s - \sigma(\vartheta)\}$ at $\vartheta = \hat{\vartheta}$ yields

$$\tilde{\Delta}' V\sqrt{n}(s - \tilde{\sigma}) = \hat{\Delta}'\hat{V}\hat{\Delta}\sqrt{n}(\hat{\vartheta} - \tilde{\vartheta}) + o_p(1), \qquad \textbf{(A.7)}$$

because Δ is continuously differentiable in ϑ and $\hat{\vartheta}$ and $\tilde{\vartheta}$ are consistent. Premultiplying both sides of (A.7) by $J(\hat{\Delta}'\hat{V}\hat{\Delta})^{-1}$, and using (A.5) and the consistency of $\hat{\Delta}$ and $\hat{V}$, we obtain

$$\sqrt{n}(\hat{\tau} - \tau) = J(\Delta' V\Delta)^{-1}\Delta' V\Xi\sqrt{n}(\bar{q} - \bar{\phi}) + o_p(1); \qquad \textbf{(A.8)}$$

consequently, the asymptotic variance matrix of $\sqrt{n}(\tau - \hat{\tau})$ is

$$\begin{aligned}\text{avar}(\hat{\tau}) &= \frac{1}{n} J(\Delta' V\Delta)^{-1}\Delta' V\Xi\Gamma^*_{\bar{q}}\Xi' V\Delta(\Delta' V\Delta)^{-1}J' \\ &= J(\Delta' V\Delta)^{-1}\Delta' V\Gamma^*_s V\Delta(\Delta' V\Delta)^{-1}J',\end{aligned}$$

as a result of (24) and equality (26). This proves statement 1 of Theorem 1.

Because Δ is in the column space of Γ_s, $\Delta' V^*\Gamma^*_s V^*\Delta = \Delta' V^*\Delta$ (see (15)). This proves the first part of statement 2.

Because for any nonnegative definite matrix V,

$$(\Delta' V\Delta)^{-1}\Delta' V\Omega V\Delta(\Delta' V\Delta)^{-1} - (\Delta'\Omega^{-1}\Delta)^{-1}$$

is positive semidefinite, so is

$$\text{avar}(\hat{\tau}) - \frac{1}{n} J(\Delta'\Omega^{-1}\Delta)^{-1}J', \qquad \textbf{(A.9)}$$

and this difference vanishes only when $V = \Omega^{-1}$. Hence the asymptotic efficiency in statement 2.

Using (A.3), the variance matrix $\text{avar}(\hat{\vartheta})$ of (10) decomposes to

$$(\Delta' V\Delta)^{-1}\Delta' V\Xi\Gamma^*_{\bar{q}}\Xi' V\Delta(\Delta' V\Delta)^{-1} + \tilde{J}\Gamma_{\bar{q}_u}\tilde{J}', \qquad \textbf{(A.10)}$$

where $\tilde{J} = (\Delta' V \Delta)^{-1} \Delta' V \Xi_u$. Because Ξ_u is a submatrix of Δ, $\tilde{J}$ is a matrix with elements 0 and 1. The first term in (A.10) is the same as under NT, and the second is free of the estimation method used; consequently, minimum variance is attained for the same estimator as under NT, i.e., when $V = V^*$. This proves statement 3. For proving the remainder of Theorem 1 we need the following lemma.

LEMMA 2. *Let* $\hat{P} = I - \hat{\Delta}(\hat{\Delta}'\hat{V}\hat{\Delta})^{-1}\hat{\Delta}'\hat{V}$ *and* $P = I - \Delta(\Delta' V \Delta)^{-1}\Delta' V$. *Under Assumptions A.2 and A.1,*

$$\hat{P}\sqrt{n}(s - \hat{\sigma}) = P\Xi\sqrt{n}(\bar{q} - \bar{\phi}) + o_p(1). \tag{A.11}$$

Proof. First,

$$\hat{P}\sqrt{n}(s - \hat{\sigma}) = \hat{P}\sqrt{n}(s - \tilde{\sigma}) + \hat{P}\sqrt{n}(\tilde{\sigma} - \dot{\sigma}) + \hat{P}\sqrt{n}(\dot{\sigma} - \hat{\sigma}), \tag{A.12}$$

where $\dot{\sigma} = \sigma(\dot{\vartheta})$ and $\dot{\vartheta} = (\hat{\tau}', \tilde{q}_u')'$. By Taylor series expansion we obtain

$$\hat{P}\sqrt{n}(\tilde{\sigma} - \dot{\sigma}) = P\Delta\sqrt{n}(\tilde{\vartheta} - \dot{\vartheta}) + o_p(1) = o_p(1),$$

because $\sqrt{n}(\tilde{\vartheta} - \dot{\vartheta}) = O_p(1)$ and $\hat{P}\tilde{\Delta} \xrightarrow{P} P\Delta = 0$. As a result of $\hat{P}\sqrt{n}(\dot{\sigma} - \hat{\sigma}) = \hat{P}\hat{\Xi}_u\sqrt{n}(\hat{\psi} - \tilde{q}_u) = 0$, (A.12) simplifies to

$$\hat{P}\sqrt{n}(s - \hat{\sigma}) = \hat{P}\sqrt{n}(s - \tilde{\sigma}) + o_p(1) = P\Xi\sqrt{n}(\bar{q} - \bar{\phi}) + o_p(1)$$

(see (20) and (21)). ■

Statement 4 follows directly from Lemma 2 and the asymptotic normality of $\bar{q}$ in Lemma 1.

Premultiplying (A.11) by $\hat{\Delta}_\perp'$ and noting that $\hat{\Delta}_\perp' \hat{P} = \hat{\Delta}_\perp'$, we obtain by Lemma 1 the asymptotic normality of $\hat{\Delta}_\perp'\sqrt{n}(s - \hat{\sigma})$; its asymptotic variance matrix is

$$\Delta_\perp' \Xi \Gamma_{\bar{q}}^* \Xi' \Delta_\perp = \Delta_\perp' \Gamma_s^* \Delta_\perp,$$

as a result of (24). Thus, by the standard theory of Wald type test statistics (e.g., Moore, 1977; Newey, 1985), the statistic

$$n(s - \hat{\sigma})'\hat{\Delta}_\perp(\hat{\Delta}_\perp'\hat{\Gamma}^*\hat{\Delta}_\perp)^{+}\hat{\Delta}_\perp'(s - \hat{\sigma}) \tag{A.13}$$

has the asymptotic chi-square distribution stated in statement 5. The proof concludes by noting that (A.13) is numerically equal to T_V^*, because $\hat{\Delta}$ and $(s - \hat{\sigma})$ are in the column space of $\hat{\Gamma}_s^*$. To prove (27), the noncentral case, we just need to combine the standard theory of Moore (1977) with the obvious result that $\sqrt{n}(s - \hat{\sigma})$ now has asymptotic mean δ, and the fact that Γ is not affected by the sample cross-product moments involving ν_{gi} (as these sample moments are terms of order $O_p(1/\sqrt{n})$). ■